# Bootstrap Approximation to Prediction MSE for State-Space Models with Estimated Parameters

**Danny Pfeffermann, Richard Tiller**

## Abstract

We propose a simple but general bootstrap method for estimating the Prediction Mean Square Error (PMSE) of the state vector predictors when the unknown model parameters are estimated from the observed series. As is well known, substituting the model parameters by the sample estimates in the theoretical PMSE expression that assumes known parameter values results in under-estimation of the true PMSE. Methods proposed in the literature to deal with this problem in state-space modelling are inadequate and may not even be operational when fitting complex models, or when some of the parameters are close to their boundary values. The proposed method consists of generating a large number of series from the model fitted to the original observations, re-estimating the model parameters using the same method as used for the observed series and then estimating separately the component of PMSE resulting from filter uncertainty and the component resulting from parameter uncertainty. Application of the method to a model fitted to sample estimates of employment ratios in the U.S.A. that contains eighteen unknown parameters estimated by a three-step procedure yields accurate results. The procedure is applicable to mixed linear models that can be cast into state-space form. (Revised on 6th Oct 2004)

**S$^3$RI Methodology Working Paper M03/05**

# Bootstrap Approximation to Prediction MSE for State-Space Models with Estimated Parameters

Danny Pfeffermann, *Hebrew University and University of Southampton*

and

Richard Tiller*, Bureau of Labor Statistics, Washington, DC.*

**Mailing address:**

Danny Pfeffermann,
Department of Statistics
Hebrew University
Jerusalem, 91905
Israel

Email: msdanny@mscc.huji.ac.il

# Bootstrap Approximation to Prediction MSE for State-Space Models with Estimated Parameters


## Abstract

We propose simple parametric and nonparametric bootstrap methods for estimating the prediction mean square error (PMSE) of state vector predictors that use estimated model parameters. As is well known, substituting the model parameters by their estimates in the theoretical PMSE expression that assumes known parameter values results in under-estimation of the true PMSE. The parametric method consists of generating parametrically a large number of bootstrap series from the model fitted to the original series, re-estimating the model parameters for each series using the same method as used for the original series and then estimating the separate components of the PMSE. The nonparametric method generates the series by bootstrapping the standardized innovations estimated for the original series. The bootstrap methods are compared to other methods considered in the literature in a simulation study that also examines the robustness of the various methods to non-normality of the model error terms. Application of the bootstrap method to a model fitted to employment ratios in the U.S.A. that contains eighteen unknown parameters, estimated by a three-step procedure yields unbiased PMSE estimators.



KEY WORDS: Hyper-parameters; Kalman filter; MLE, Order of bias, REML

## 1. INTRODUCTION

The state-space model considered in this article consists of two sets of equations;

1- *The observation (measurement) equation*:

$$y_t = Z_t u_t + \varepsilon_t \quad ; \quad \varepsilon_t \sim N(0, \Sigma_t), \; E(\varepsilon_t \varepsilon'_{t-k}) = 0, \quad k > 0 \tag{1.1}$$

2- *The state (transition) equation*:

$$u_t = G_t u_{t-1} + \eta_t \quad ; \quad \eta_t \sim N(0, Q_t), \; E(\eta_t \eta'_{t-k}) = 0, \quad k > 0 \tag{1.2}$$

It is assumed also that $E(\varepsilon_t \eta'_s) = 0$ for all $t$ and $s$. Note that both $y_i$ and $u_t$ can be vectors. Although not written in the most general form, the state-space model defined by (1.1) and (1.2) is known to include as special cases many of the time series models in common use, see Harvey (1989) for illustrations. Auxiliary variables can be added to both equations. It is natural to think of state-space models as time series models but it is important to note that familiar mixed linear models can also be cast into state-space form so that the bootstrap method proposed in this article for the estimation of the PMSE applies to these models as well. See, e.g., Sallas and Harville (1981) for the presentation of mixed linear models in state-space form.

When fitting state-space models, the focus of the analysis is ordinarily the prediction of linear functions $\alpha_t = l'_t u_t$ of the components of the state vector with known coefficients $l_t$. Simple examples are the prediction of future values of the series $y_t$, or the prediction of the trend and seasonally adjusted values in the Basic Structural Model (Harvey, 1989; section 4). A major advantage of the state-space model representation is that the predictor of the state vector for any given time $t$ based on observations $y_1 ... y_n$ is easily obtained by means of the Kalman filter for $t \geq n$, or by an appropriate smoothing algorithm for $t < n$. Moreover, when all the model 'hyper-parameters' are known, the use of these filters also yields the corresponding prediction MSE (PMSE). By the model hyper-parameters we mean the elements of the covariance matrices $\Sigma_t$, $Q_t$ and possibly also elements of the matrices $Z_t$ or $G_t$. (The Kalman filter equations are shown in Appendix A, see Harvey, 1989 and de Jong, 1989 for smoothing algorithms.)

In actual applications the model hyper-parameters are seldom known. A common practice is to estimate them and substitute the sample estimates in the theoretical expressions of the state predictors and the PMSE. The use of this practice may result, however, in severe underestimation of the true PMSE, particularly with short series, as the resulting MSE estimators ignore the variability implied by the parameter estimation. A similar problem arising in Small Area Estimation evoked extensive research in the last two decades on plausible bias corrections; see Pfeffermann (2002) for a recent review.

The purpose of this article is to develop simple parametric and nonparametric bootstrap procedures for the computation of valid PMSE estimators in the practical situations where the state vector predictors use estimated hyper-parameter values. We follow the frequentist approach by which the true hyper-parameters are considered fixed and the PMSE is evaluated over the joint distribution of the state vectors and the measured values. The parametric procedure consists of generating parametrically a large number of bootstrap series from the model fitted to the original series, re-estimating the model hyper-parameters for each series using the same method as used for the observed series and then estimating the separate components of the PMSE. The nonparametric procedure generates the series by bootstrapping the standardized innovations estimated for the original series. Bootstrapping of state-space models has been considered before by Stoffer and Wall (1991, 2002), but these studies address different problems (see section 3). The need of developing valid PMSE estimators for state-space models is often raised by researchers working in this field, see, e.g., Durbin and Koopman (2000) and the discussion of A. Harvey to that article.

Section 2 contains a more rigorous discussion of the problem of estimating the PMSE of state vector estimators that use estimated hyper-parameters. Section 3 describes the bootstrap procedures and compares their properties to other methods proposed in the literature to deal with the problem. The various methods are further compared in Section

4 by means of a simulation study that also examines the robustness of the various methods to non-normality of the model error terms. The performance of the bootstrap method is further examined by applying it to a model fitted to employment ratios in the U.S.A. that contains eighteen unknown hyper-parameters. This model is similar to the models used by the Bureau of Labor Statistics in the U.S.A. for the production of employment and unemployment State estimates. Section 5 contains a brief summary with possible applications of the method to different state-space models.

## 2. STATEMENT OF PROBLEM

In what follows we consider the model defined by (1.1) and (1.2) and focus on the prediction of functions $\alpha_t = l_t' u_t$ of the state vector. In Section 4 we consider other distributions for the model error terms $\varepsilon_t$ and $\eta_t$. Let $y_{(n)} = \langle y_1 ... y_n \rangle$ represent the observed series and denote by $\lambda$ the vector of model hyper-parameters contained in $\Sigma_t, Q_t$, and possibly also in $Z_t$ and $G_t$. For known $\lambda$, the optimal state predictor and the corresponding PMSE are defined as,

$$\alpha_t(\lambda) = E[\alpha_t \mid y_{(n)}; \lambda] \quad ; \qquad P_t(\lambda) = E\{[\alpha_t - \alpha_t(\lambda)]^2 \mid y_{(n)}; \lambda\} \tag{2.1}$$

The predictor $\alpha_t(\lambda)$ is the posterior mean of $\alpha_t$ under the Bayesian approach, and is the best predictor (minimum MSE) under the Frequentist approach. It is the best linear unbiased predictor (BLUP) of $\alpha_t$ when relaxing the normality assumption for the error terms, with $P_t(\lambda)$ defining the PMSE in all the three cases. Notice again that *t* may be smaller, equal or larger than *n* and that the predictor and PMSE in (2.1) can be obtained by application of the Kalman filter or an appropriate smoothing algorithm, utilizing the relationship

$$f(\alpha_t \mid y_{(n)}; \lambda) \prec f(y_n \mid \alpha_t, y_{(n-1)}; \lambda) f(\alpha_t \mid y_{(n-1)}; \lambda) \tag{2.2}$$

with $\prec$ denoting proportionality.

Our interest in this article is in the practical case where $\lambda$ is replaced by sample estimates in the expression for the state predictor and the problem considered is how to evaluate the corresponding PMSE. A Bayesian solution to the problem consists of specifying a prior distribution for $\lambda$ and computing the expectation of $P_t(\lambda)$ in (2.1) over the posterior distribution of $\lambda$. See Section 3.4 In the rest of this paper we follow the frequentist approach by which $\lambda$ is considered fixed and the PMSE is evaluated with respect to the joint distribution of the state vectors and the *y*-values.

Denote by $\hat{\lambda}$ the vector of hyper-parameter estimates and by $\alpha_t(\hat{\lambda})$ the predictor obtained from $\alpha_t(\lambda)$ defined in (2.1) by substituting $\hat{\lambda}$ for $\lambda$. The prediction error is in this case, $[\alpha_t(\hat{\lambda})-\alpha_t]=[\alpha_t(\hat{\lambda})-\alpha_t(\lambda)]+[\alpha_t(\lambda)-\alpha_t]$ and the PMSE is,

$$MSE_t = E[\alpha_t(\hat{\lambda})-\alpha_t]^2 = E[\alpha_t(\lambda)-\alpha_t]^2 + E[\alpha_t(\hat{\lambda})-\alpha_t(\lambda)]^2 \tag{2.3}$$

The expectations in (2.3) are over the joint distribution of $\alpha_t$ and $y_{(n)}$, as defined by (1.1) and (1.2). Notice that,

$$E\{[\alpha_t(\hat{\lambda})-\alpha_t(\lambda)][\alpha_t(\lambda)-\alpha_t]\} = E_{y_{(n)}}\{E_{\alpha_t|y_{(n)}}[\alpha_t(\hat{\lambda})-\alpha_t(\lambda)][\alpha_t(\lambda)-\alpha_t] \mid y_{(n)}\} = 0 \tag{2.4}$$

since $[\alpha_t(\hat{\lambda})-\alpha_t(\lambda)]$ is fixed when conditioning on $y_{(n)}$ and under normality of the error terms, $\alpha_t(\lambda)=E[\alpha_t \mid y_{(n)}]$.

The PMSE in (2.3) is factorized into two components. The first component, $P_t(\lambda)=E[\alpha_t(\lambda)-\alpha_t]^2$ is the contribution to the PMSE resulting from 'filter uncertainty'. This is the true PMSE if $\lambda$ were known (compare with 2.1, for known $\lambda$ the PMSE does not depend on the observed series). The second component, $E[\alpha_t(\hat{\lambda})-\alpha_t(\lambda)]^2$ is the contribution to the PMSE resulting from 'parameter uncertainty'. For $\hat{\lambda}$ such that $E[(\hat{\lambda}-\lambda)(\hat{\lambda}-\lambda)']=O(1/n)$ and under some other regularity conditions, this component is of order $O(1/n)$. This property can be shown to hold for ARMA models using results from Ansley and Kohn (1986). Conditions guaranteeing this order for the general state-space model defined by (1.1) and (1.2) are stated in Appendix B.

Next consider the 'naïve' PMSE estimator $P_t(\hat{\lambda})$, obtained by substituting $\hat{\lambda}$ for $\lambda$ in (2.1). The use of this estimator ignores the second component on the right hand side of (2.3). Furthermore, for $\hat{\lambda}$ such that $E(\hat{\lambda}-\lambda)=O(1/n)$ and $E[(\hat{\lambda}-\lambda)(\hat{\lambda}-\lambda)']=O(1/n)$, $E[P_t(\hat{\lambda})-P_t(\lambda)]=O(1/n)$, the same order as the order of the neglected component $E[\alpha_t(\hat{\lambda})-\alpha_t(\lambda)]^2$. This follows by expanding $P_t(\hat{\lambda})$ around $P_t(\lambda)$ and assuming that the derivatives of $P_t(\lambda)$ with respect to $\lambda$ are bounded. For models that are time invariant such that the matrices $\Sigma_t, Q_t, Z_t, G_t$ are fixed over time, and under some regularity conditions, $\lim_{t\to\infty} P_t(\lambda)=\overline{P}<\infty$, see Harvey (1989) for details.

The method proposed in the next section for estimating the PMSE accounts for both components of (2.3), with bias of order $O\left(1/n^2\right)$.

## 3. BOOTSTRAP METHODS FOR ESTIMATION OF PMSE

### 3.1 Parametric bootstrap

The method consists of three steps:

1- Generate (parametrically) a large number B of state vector series $\{u_t^b\}$ and observations $\{y_t^b\}$ $(t=1...n)$ from the model (1.1)-(1.2), with hyper parameters $\lambda=\hat{\lambda}$ estimated for the original series.

2- Re-estimate the vector of hyper-parameters for each of the generated series using the same method as used for estimating $\hat{\lambda}$, yielding estimates $\{\hat{\lambda}^b\}$ $(b=1...B)$.

3- Estimate $MSE_t=E[\alpha_t(\hat{\lambda})-\alpha_t]^2$ as,

$$M\hat{S}E_t = MSE_{t,p}^{bs} + 2P_t(\hat{\lambda}) - \overline{P}_t^{bs} \tag{3.1}$$

where,

$$MSE_{t,p}^{bs} = \frac{1}{B}\sum_{b=1}^{B}[\alpha_t^b(\hat{\lambda}^b) - \alpha_t^b(\hat{\lambda})]^2 \;\; ; \;\; \bar{P}_t^{bs} = \frac{1}{B}\sum_{b=1}^{B} P_t(\hat{\lambda}^b). \qquad (3.2)$$

In (3.2), $\alpha_t^b(\hat{\lambda}) = l_t' u_t^b(\hat{\lambda})$ and $\alpha_t^b(\hat{\lambda}^b) = l_t' u_t^b(\hat{\lambda}^b)$, with $u_t^b(\hat{\lambda})$ and $u_t^b(\hat{\lambda}^b)$ defining the state predictors that use hyper-parameter estimates $\hat{\lambda}$ and $\hat{\lambda}^b$ respectively. The symbol $P_t(\hat{\lambda}^b)$ defines the *(naive)* PMSE estimator that uses $\hat{\lambda}^b$.

The following theorem is proved in Appendix C where the expectation E is with respect to the joint distribution of the state vectors and the *y*-values.

*Theorem:* Let $\hat{\lambda}$ be such that $E(\hat{\lambda} - \lambda) = O(1/n)$ and $E[(\hat{\lambda} - \lambda)(\hat{\lambda} - \lambda)'] = O(1/n)$. If,

I - $E[P_t(\hat{\lambda}) - P_t(\lambda)] = O(1/n)$

II - $E[\alpha_t(\hat{\lambda}) - \alpha_t(\lambda)]^2 = O(1/n)$

Then, as $B \to \infty$, $E[M\hat{S}E_t - MSE_t] = O(1/n^2)$.

As mentioned before, the rate in I holds under very mild conditions. Conditions guaranteeing the rate in II are given in Appendix B.

The estimator $M\hat{S}E_t$ is the sum of two estimators:

1- Estimator of 'filter uncertainty'; $\hat{P}_t(\lambda) = \hat{E}[\alpha_t(\lambda) - \alpha_t]^2 = 2P_t(\hat{\lambda}) - \bar{P}_t^{bs} \to$ resembles the familiar bootstrap bias correction,

2- Estimator of 'parameter uncertainty'; $\hat{E}[\alpha_t(\hat{\lambda}) - \alpha_t(\lambda)]^2 = \frac{1}{B}\sum_{b=1}^{B}[\alpha_t^b(\hat{\lambda}^b) - \alpha_t^b(\hat{\lambda})]^2 \to$ the bootstrap analogue of $E[\alpha_t(\hat{\lambda}) - \alpha_t(\lambda)]^2$.

An alternative estimator of $MSE_t$ (same order, see the proof of the theorem) is,

$$M\hat{S}E_t = P_t(\hat{\lambda}) - \bar{P}_t^{bs} + \frac{1}{B}\sum_{b=1}^{B}[\alpha_t^b(\hat{\lambda}^b) - \alpha_t^b]^2 \tag{3.3}$$

where $\alpha_t^b = l_t' u_t^b$. Note that $M\hat{S}E_t^{bs} = \frac{1}{B}\sum_{b=1}^{B}[\alpha_t^b(\hat{\lambda}^b) - \alpha_t^b]^2$ is the bootstrap PMSE but as implied by the proof of the theorem, the use of this term alone is not sufficient for estimating the PMSE with bias of correct order.

Generating series from the Gaussian model (1.1)-(1.2) with given (estimated) hyper-parameters is straightforward. Basically, what is required is to generate independent error vectors $\varepsilon_t$ and $\eta_t$ from the corresponding normal distributions (or other distributions underlying the model), generate the state vectors $u_t^b$ using (1.2) with an appropriate initialization, and then generate the series $y_t^b$ using (1.1).

### 3.2 Nonparametric bootstrap

This method differs from the parametric bootstrap method in the way that the bootstrap series are generated. This is done by repeating the following 2 steps B times.

*Step* 1: Express the model for the states $u_t$ and the measurements $y_t$ as a function of the model innovations (one step ahead prediction errors), $v_t = y_t - \hat{y}_{t|t-1} = y_t - Z_t u_{t|t-1}(\lambda)$, where $u_{t|t-1}(\lambda) = G_t u_{t-1}(\lambda)$ is the predictor at time *t-1* of the state vector at time *t* (Equation A2 in Appendix A). Compute the empirical innovations and the corresponding variances by application of the Kalman filter, with the hyper-parameters $\lambda$ set at their estimated values, $\hat{\lambda}$. Compute the empirical standardized innovations (Equation A3 in Appendix A with $\lambda$ replaced by $\hat{\lambda}$).

*Step* 2. Sample with replacement *n* standardized innovations from the standardized innovations computed in Stage 1 and construct a bootstrap series of observations using the relationships in Equation A2 of Appendix A after an appropriate initialization. Re-estimate the hyper-parameters $\lambda$.

The MSE estimators are obtained under the nonparametric method in the same way as under the parametric method, using Equations (3.1) and (3.2).

It should be noted that the nonparametric bootstrap method is not completely 'model free'. This is so for two reasons. First, the common use of maximum likelihood estimation (MLE) for the hyper-parameters requires distributional assumptions. Second, the use of the estimator defined by (3.1)-(3.2) assumes the decomposition (2.3) of the MSE, or the zeroing of the cross-product expectation in (2.4), which is not necessarily true under non-normal distributions of the error terms. Notice also in this regard that the bootstrap estimator defined by (3.3) is not operational under the nonparametric method. Stoffer and Wall (1991) use nonparametric bootstrapping of the empirical standardized innovations for estimating the distribution of the MLE, $\hat{\lambda}$. Stoffer and Wall (2002) use a similar bootstrap method for estimating the conditional distribution of the forecasts of the *y*-series given the last observation. These two problems are different from the problem considered in the present article. In particular, the authors estimate both distributions by the corresponding bootstrap distributions but as emphasized below (3.3), the PMSE of the state predictors cannot be estimated by the bootstrap PMSE alone, with bias of correct order.

An interesting question underlying the use of the bootstrap method is the actual number of bootstrap samples that need to be generated. In the simulation study described in Section 4 with a complex model that contains 18 unknown parameters and series of length 84, the use of 500 series was found to yield unbiased PMSE estimators, but this outcome doesn't necessarily generalize to other models and series lengths. The determination of the number of bootstrap samples is not a trivial problem. See Shao and Tu (1995) for discussion and guidelines with references to other studies.

### 3.3 Other Methods Proposed in the Literature

The problem considered in this article had been studied previously. Ansley and Kohn (1986) propose to approximate $P_t(\lambda)$, (the first term in (2.3)) by $P_t(\hat{\lambda})$ and expand $\alpha_t(\hat{\lambda})$ around $\alpha_t(\lambda)$ for approximating the second term. The resulting PMSE estimator is,

$$M\hat{S}E_t^{AK} = P_t(\hat{\lambda}) + \{[\frac{\partial \alpha_t(\lambda)}{\partial \lambda}]_{\lambda=\hat{\lambda}}\}'[\mathrm{I}^{-1}(\hat{\lambda})]\{[\frac{\partial \alpha_t(\lambda)}{\partial \lambda}]_{\lambda=\hat{\lambda}}\} \quad (3.4)$$

where $\hat{\lambda}$ is the MLE of $\lambda$ and $\mathrm{I}(\hat{\lambda})$ is the corresponding information matrix evaluated at $\hat{\lambda}$. The estimator (3.4) is derived from a frequentist standpoint but as noted by the authors, it also has a Bayesian interpretation. Under the ('empirical') Bayesian approach, the true PMSE when estimating the hyper-parameters is computed as,

$$MSE_{t|y_{(n)}} = E_{\pi[\lambda|y_{(n)}]}[P_t(\lambda)] + E_{\pi[\lambda|y_{(n)}]}[\alpha_t(\lambda) - \alpha_t(\hat{\lambda})]^2 \quad (3.5)$$

where $\pi[\lambda \mid y_{(n)}]$ is the posterior distribution of $\lambda$. Thus, the estimator (3.4) can be interpreted as an approximation to (3.5) under the assumption that $\pi[\lambda \mid y_{(n)}]$ can be approximated by the normal distribution $N[\hat{\lambda}, I^{-1}(\hat{\lambda})]$.

Hamilton (1986), following the Bayesian perspective above, proposes to generate a large number $M$ of realizations $\lambda_1 ... \lambda_M$ from the posterior $\pi[\lambda \mid y_{(n)}]$ and estimate the PMSE in (3.5) as,

$$M\hat{S}E_t^H = \frac{1}{M}\sum_{i=1}^M P_t(\lambda_i) + \frac{1}{M}\sum_{i=1}^M [\alpha_t(\lambda_i) - \alpha_t(\hat{\lambda})]^2 \quad (3.6)$$

The posterior $\pi[\lambda \mid y_{(n)}]$ is again approximated by the normal distribution, $N[\hat{\lambda}, I^{-1}(\hat{\lambda})]$.

In a recent article, Quenneville and Singh (2000) show that estimating $E_{\pi[\lambda|y_{(n)}]}[P_t(\lambda)]$ (the first term of 3.5) by $\frac{1}{M}\sum_{i=1}^M P_t(\lambda_i)$ as in Hamilton (1986), or by $P_t(\hat{\lambda})$ as in Ansley and Kohn (1986) yields in both cases a bias of order O(1/n). The authors propose therefore enhancements to reduce the order of the bias, which consist of replacing the MLE $\hat{\lambda}$ by an estimator $\bar{\lambda}$ satisfying $E(\bar{\lambda} \mid \lambda) = \lambda + O(n^{-2})$. (The enhancement to Ansley and Kohn approach also involves adding a term of order $O(1/n)$.) The estimator $\bar{\lambda}$ is obtained by maximizing a modification of the likelihood equations used for the computation of the restricted MLE.

The use of the above procedures for bias correction has four disadvantages.

1- The original PMSE estimators of Ansley and Kohn (1986) and Hamilton (1986) have bias of order *O*(*1/n*), which is the order of the PMSE (see below Equation 2.4). As explained by Quenneville and Singh (2000), the estimator of the PMSE needs to be unbiased up to terms of order smaller than *O(1/n)*.

2- All the methods approximate the posterior $\pi[\lambda \mid y_{(n)}]$ of the hyper-parameters by the normal distribution, which is not always justified. This approximation stems from the asymptotic normality property of the MLE, but the distribution of the MLE can be skewed, particularly for short series, or when some of the hyper-parameters are close to their boundary values. Transformation of the hyper-parameters often improves the approximation, but only to a certain extent. Figure 1 shows the empirical distribution of the MLE of the logs of the state variance estimator, $\hat{x} = 0.5 log(\hat{q})$, for series of length 40 generated from the simple Gaussian 'random walk plus noise model'. (See Section 4.1 for more details. The use of this transformation for model variances is very common, see e.g., Koopman *et. al.* 1995, Page 210). Testing the normality of this distribution yields p-values < 0.01 with all the common normality tests. The mean is -1.20, the median is -0.80 and the skewness is -5.04. The distribution of $\hat{x}$ is much closer to normality when increasing the length of the series to 100, but even in this case the normality hypothesis is rejected by all the tests, with p-values < 0.01.

Figure 2 shows the empirical distribution of the logs of the MLE of the slope variance $\sigma_R^2$ under the BSM model defined by (4.4) for series of length 84, with true variance $\sigma_R^2 = 0.0024$. As readily seen, the distribution is very skewed with mean, median and skewness equal to -9.68, -6.41 and -2.19 respectively, which in this case is explained by the proximity of the true variance to its boundary value. A similar picture is obtained even when increasing the length of the series to 240.

Stoffer and Wall (1991) likewise discuss the limitations of assuming normality for MLE in state-space modelling.

3- The computation of the Information matrix required for these methods may become unstable as the model becomes more complex and the number of unknown parameters increases. See Quenneville and Singh (200) for further discussion.

4- As already implied by the preceding discussion, all these methods basically assume that the model hyper-parameters are estimated by MLE or REML. This is not necessarily the case in practice and at least some of the parameters could be estimated by different methods, possibly using different data sources.

The use of the bootstrap methods overcomes the four disadvantages mentioned above. In particular, it produces estimators with bias of order $O(1/n^2)$, it does not rely on normality of the hyper-parameter estimators and is not restricted to MLE or REML estimators of the hyper-parameters (see Section 4.2). The empirical results in Section 4.1 further support the use of these methods.

### 3.4 The Full Bayesian Approach

In the (empirical) Bayesian method mentioned in Section 3.3 the unknown hyper-parameters are first estimated by MLE or REML and then the PMSE is evaluated by computing the expectation of $\{P_t(\lambda)+[\alpha_t-\alpha_t(\hat{\lambda})]^2\}$ over the posterior distribution $\pi[\lambda \mid y_{(n)}]$, see Equation (3.5). An alternative method of predicting the state vector and evaluating the PMSE is the use of the full Bayesian paradigm. The Bayesian solution consists of specifying a prior distribution $\pi(\lambda)$ for $\lambda$ and integrating the PMSE in (2.1) with respect to the posterior distribution $\pi[\lambda \mid y_{(n)}]$ yielding,

$$\begin{aligned} \alpha_t{*} &= E_{\pi[\lambda \mid y_{(n)}]}\{E(\alpha_t \mid y_{(n)},\lambda)\} = E[\alpha_t \mid y_{(n)}] \\ MSE_t{*} &= E_{\pi[\lambda \mid y_{(n)}]}\{E[(\alpha_t-\alpha_t{*})^2 \mid y_{(n)},\lambda]\} = E[(\alpha_t-\alpha_t{*})^2 \mid y_{(n)}] \end{aligned} \tag{3.7}$$

The major advantage of this approach is that it yields the posterior variance of the predictors, accounting for all sources of variation. On the other hand, it requires the specification of a prior distribution for $\lambda$, and for complex models the computations become very heavy and time consuming even with modern computing technology. See, for example, the article by Datta *et al.* (1999) for a recent implementation of the Markov Chain Monte Carlo (MCMC) approach with a model fitted to unemployment rates in the 50 States of the U.S.A.

## 4. EMPIRICAL RESULTS

### 4.1 Comparison of methods

This section compares the bootstrap methods with the methods discussed in Section 3 by repeating the simulation study performed by Quenneville and Singh (2000). The experiment consists of generating S=1000 series from the 'random walk plus noise' (RWN) model and estimating the PMSE of the empirical predictor $\hat{\alpha}_t$ for every time point *t* by each of the methods. The RWN model is defined as,

$$y_t = u_t + \varepsilon_t \ , \ u_t = u_{t-1} + \eta_t \quad ; \quad \varepsilon_t \sim N(0,\sigma^2) \ , \ \eta_t \sim N(0,\sigma^2 q) \qquad (4.1)$$

where $\varepsilon_t$ and $\eta_t$ are mutually and serially independent. For the present experiment, $\sigma^2 = 1, \ q = 0.25$. The state value of interest is $\alpha_t = u_t$. Notice that the optimal predictor of $\alpha_t$ under the model with known variances does not depend in this case on $\sigma^2$. The empirical predictor $\hat{\alpha}_t$ is obtained by replacing *q* by its REML estimator $\hat{q}$ in the expression of the optimal predictor, where $\hat{q}$ is calculated by first calculating the REML of $x = 0.5log(q)$ and then defining $\hat{q} = exp(2\hat{x})$. The restricted log likelihood equations for state space models are developed in Tsimikas and Ledolter (1994, Equation 2.13). The REML $\hat{\sigma}^2$ required for the computation of the MSE estimators is available analytically as a function of $\hat{q}$, (the variance $\sigma^2$ is concentrated out of the likelihood). We considered series of length 40 and 100 and computed the true PMSE of $\hat{\alpha}_t$ for given *t* by simulating 50,000 series for each length; $MSE_t = \sum_{i=1}^{50,000} (\hat{\alpha}_{t,i} - \alpha_{t,i})^2 / 50,000$, $t = 1 ... T$ $(T = 40,100)$. For other computational details underlying this experiment see the article by Quenneville and Singh (2000).

Table 1 shows the mean percent relative bias (Rel-Bias) and mean percent relative root mean square error (Rel-RMSE) of the following MSE estimators: Na?ve (N), obtained by substituting $(\hat{\sigma}_u^2, \hat{q})$ for $(\sigma_u^2, q)$ in the expression for the PMSE of the optimal predictor $\alpha_t(q)$ that employs the true variance; Ansley and Kohn (AK), defined by (3.4); Hamilton (H), defined by (3.6); The corrected estimators of Ansley and Kohn ($AK_c$), and Hamilton ($H_c$) developed by Quenneville and Singh (2000); parametric bootstrap (PB) developed in Section 3.1 (Equation 3.1) and nonparametric bootstrap (NPB) described in Section 3.2. The results for the na?ve estimator, AK and $AK_c$ are based on 5000 simulated series. The results for H, $H_c$ and the bootstrap methods are based on 1000 simulated series, drawing 2000 values $q_i$ for H and $H_c$ and generating 2000 bootstrap series for each simulated series for the bootstrap methods.

Denote by $d_{s,t} = [M\hat{S}E_{s,t} - MSE_t]$ the error in estimating the true MSE at time *t* with series *s* by an estimator $M\hat{S}E_{s,t}$ and let $\bar{d}_t = \sum_{s=1}^{1000} d_{s,t}/1000$, $\bar{d}_t^{(2)} = \sum_{s=1}^{1000} d_{s,t}^2/1000$. The mean Rel-Bias and Rel-RMSE are defined as,

$$\text{Rel-Bias} = \frac{100}{T}\sum_{t=1}^{T}[\bar{d}_t / MSE_t] \quad , \quad \text{Rel-RMSE} = \frac{100}{T}\sum_{t=1}^{T}[(\bar{d}_t^{(2)})^{1/2} / MSE_t] \qquad (4.2)$$

**Table 1. Percent Mean Relative Bias and Relative Root MSE of PMSE Estimators for the RWN Model with Normal Errors**

| | T=40 | | T=100 | |
|---|---|---|---|---|
| Method | Rel-Bias | Rel-RMSE | Rel-Bias | Rel-RMSE |
| N | -18.50 | 33.74 | -7.56 | 18.41 |
| H | -9.51 | 36.65 | -4.34 | 18.76 |
| AK | -7.52 | 36.95 | -3.19 | 18.67 |
| $H_c$ | -10.66 | 36.23 | -2.63 | 18.89 |
| $AK_c$ | -7.62 | 37.45 | -2.40 | 18.47 |
| PB | 0.63 | 34.11 | 1.59 | 17.03 |
| NPB | -1.09 | 34.14 | 0.55 | 18.56 |

The first 6 estimators in Table 1 have been considered by Quenneville and Singh (2000) and even though we attempted to emulate their experiment exactly, the biases obtained in our study are always substantially lower than the biases reported in their article, including for the new methods $AK_c$ and $H_c$ developed by them. However, the ordering of the methods with respect to the magnitude of the bias is the same in both studies. (Quenneville and Singh do not report the MSE of the PMSE estimators).

The results in Table 1 show that the bootstrap methods are much superior to the other methods in terms of the bias, notably with the shorter series. The two bootstrap estimators also have lower RMSEs than the other methods proposed for reducing the bias of the naïve estimator by about 6-8%, except for the long series where the use of NPB yields a similar RMSE to that of the other methods. Notice that the enhanced methods $H_c$ and $AK_c$ proposed by Quenneville and Singh (2000) indeed reduce the bias for the case $T = 100$ (but not for $T = 40$), but the Rel-RMSE are similar to the values obtained without the correction terms. An interesting outcome revealed from the table is that the naïve estimator, although being extremely biased, has similar RMSE to those of the bootstrap estimators for $T = 40$, and similar RMSE to NPB (and the other methods except PB) for $T = 100$. This result is not surprising since the addition of correction terms to control the bias increases the variance. See also Section 4.2, and Singh *et al.* (1998) for similar findings in a small area estimation study.

In order to study the sensitivity of the various methods to the normality assumptions underlying the Gaussian model, we repeated the same simulation study but this time by generating the random errors $\varepsilon_t$ and $\eta_t$ in (4.1) from Gamma distributions. Specifically,

$$\varepsilon_t = v_t - \frac{4}{3},\ v_t \sim Gamma(\frac{16}{9}, \frac{3}{4})\ ;\ \eta_t = w_t - \frac{5}{8},\ w_t \sim Gamma(\frac{25}{16}, \frac{2}{5}) \qquad (4.3)$$

Notice that the variances of the two error terms are the same as under the Gaussian model. The distribution of $\varepsilon_t$ is displayed in Figure 3 and is seen to be very skewed.

**Table 2. Percent Mean Relative Bias and Relative Root MSE of PMSE Estimators for the RWN Model with Gamma Errors**

| | T=40 | | T=100 | |
|---|---|---|---|---|
| Method | Rel-Bias | Rel-RMSE | Rel-Bias | Rel-RMSE |
| N | -18.28 | 38.54 | -7.08 | 22.19 |
| H | -8.83 | 49.43 | -4.08 | 23.35 |
| AK | -8.25 | 41.72 | -3.68 | 23.13 |
| $H_c$ | -12.11 | 42.98 | -2.52 | 23.32 |
| $AK_c$ | -8.76 | 42.78 | -2.82 | 23.11 |
| PB | -1.35 | 41.85 | 1.44 | 22.20 |
| NPB | -0.31 | 40.06 | 1.18 | 22.48 |

The biases displayed in Table 2 are quite similar to the biases in Table 1, indicating that the performance of the various methods is not sensitive to the normality assumptions. The Rel-RMSE, however, are higher in this case by about 17.5% for $T = 40$ and 25% for $T = 100$. (The increase in Rel-RMSE for Hamilton's method with $T = 40$ is 35%.) The bootstrap methods again perform better than the other methods, particularly in terms of bias reduction. Notice that like in the Gaussian case, the na?ve PMSE estimator has the lowest Rel-RMSE despite being highly biased.

### 4.2 Application of the parametric bootstrap method with a model fitted to employment rates in the U.S.A.

In this section we apply the parametric bootstrap (PB) method to the model fitted by the Bureau of Labour Statistics (BLS) in the U.S.A. to the series *Employment to Population Ratio in the District of Columbia*, abbreviated hereafter by EP-DC. The EP series consist of the percentage of employed persons out of the total population aged 15+. This is one of the key economic series in the U.S.A., published monthly by the BLS for each of the 50 States and DC. The BLS uses similar models for the production of the major employment and unemployment statistics in all the States, see Tiller (1992) for details. In order to assess the performance of the PB method, we generated a large number of series from the EP-DC model and applied the method to each of the series (by generating another set of bootstrap series). The unique feature of this experiment is that the model contains 18 unknown hyper-parameters estimated in 3 stages, with only 3 of the parameters being estimated by maximization of the likelihood.

### 4.2.1 Model and Parameter estimation

The EP-DC series is plotted in Figure 4 along with the estimated trend under the model defined below. This is a very erratic series: the irregular component (calculated by X12 ARIMA) explains 55% of the month to month changes and 32% of the yearly changes. A large portion of the irregular component is explained by the sampling errors. Let $y_t$ define the direct sample estimate at time *t* and $Y_t$ the corresponding true population ratio such that $e_t = y_t - Y_t$ is the sampling error. A state-space model is fitted to the series $y_t$ that combines a model for $Y_t$ with a model for $e_t$. The model postulated for $Y_t$ is the Basic Structural Model (BSM, Harvey, 1989),

$$\begin{aligned}
& Y_t = L_t + S_t + I_t\,, \quad I_t \sim N(0, \sigma_I^2) \\
& L_t = L_{t-1} + R_t \quad ; \quad R_t = R_{t-1} + \eta_{Rt} \quad , \quad \eta_{Rt} \sim N(0, \sigma_R^2) \\
& S_t = \textstyle\sum_{j=1}^{6} S_{j,t}\,; \\
& S_{j,t} = \cos\omega_j S_{j,t-1} + \sin\omega_j S^*_{j,t-1} + \eta_{j,t}\ ,\ \eta_{j,t} \sim N(0, \sigma_S^2) \\
& S^*_{j,t} = -\sin\omega_j S_{j,t-1} + \cos\omega_j S^*_{j,t-1} + \eta^*_{j,t}\ ,\ \eta^*_{j,t} \sim N(0, \sigma_S^2) \\
& \omega_j = 2\pi_j / 12\,,\ j = 1 \ldots 6
\end{aligned} \tag{4.4}$$

The error terms $I_t, \eta_{Rt}, \eta_{j,t}\eta^*_{j,t}$ are mutually independent normal disturbances. In (4.1) $L_t$ is the trend level, $R_t$ is the slope and $S_t$ is the seasonal effect. The model for the trend approximates a local linear trend, whereas the model for the seasonal effects uses the traditional decomposition of the seasonal component into 11 cyclical components corresponding to the 6 seasonal frequencies. The added noise enables the seasonal effects to evolve stochastically over time.

The model fitted to the sampling error is *AR*(15), which approximates the sum of an *MA*(15) process and an *AR*(2) process. The *MA*(15) process accounts for the autocorrelations implied by the sample overlap induced by the Labour Force sampling design. By this design, households in the sample are surveyed for 4 successive months, they are left out of the sample for the next 8 months and then they are surveyed again for 4 more months. The $AR(2)$ process accounts for the autocorrelations arising from the fact that households dropped from the survey are replaced by households from the same 'census tract'. These autocorrelations exist irrespective of the sample overlap. The reduced ARMA representation of the sum of the two models is ARMA(2,17), which is approximated by an AR(15) model.

The separate models holding for the population ratios and the sampling errors are cast into a single state-space model as defined by (1.1) and (1.2). Note that the state vector consists of the trend, the slope, seasonal effects and sampling errors. The monthly variances of the sampling errors are estimated externally based on a large number of replications and considered as known, implying that the combined model depends on 18 unknown hyper-parameters. In order to simplify and stabilize the maximization of the likelihood, the $AR(15)$ model coefficients are estimated externally by first estimating the autocorrelations of the sampling errors and then solving the corresponding Yule Walker equations (Box and Jenkins, 1976). The autocorrelations are estimated from the distinct monthly panel estimates as described in Pfeffermann *et al.* (1998). (A panel is defined by the group of people joining and leaving the sample in the same months. There are 8 panels in every month. The actual series is the arithmetic mean of the 8 panel series.)

Having estimated the $AR(15)$ model coefficients, the three variances of the population model (4.4) are estimated by maximization of the likelihood, with the $AR$ coefficients held fixed at their estimated values. The PB method for PMSE estimation accounts for the estimation of all the 18 unknown hyper-parameters, even though only 3 of them are estimated by maximization of the likelihood (see below).

### 4.2.2 Simulation Study

The simulation study consists of three phases. In the first phase we generated 10,000 series from the model fitted to the EP-DC series. In the second phase we predicted the trend levels, $L_t$, and the seasonally adjusted values, $y_t - S_t - e_t = L_t + I_t$, for each of the series based on newly estimated hyper-parameters, and computed the empirical PMSE of these predictors. In the third phase we applied the PB method by generating 500 bootstrap series for each of 500 series selected at random from the 10,000 primary series. All the series are of length *n=84*, same as the length of the original series.

*Phase A- Generation of primary series*

As mentioned in Section 4.2.1, the actual series is the mean of 8 separate panel series. Hence, the first step of the simulation study was to generate 10,000 primary sets of 8 streams of sampling errors from the $AR(15)$ model fitted to the original EP-DC series (see Table 3). Let $e_{tr}^{(j)}$ $(j=1...8,\ \ t=1...84)$ define the *r-th* set of stream sampling errors (*r=1…10,000*). Next we generated primary series of population values from the model (4.4), using again as hyper-parameters the variances estimated for the original EP-DC series, except for the irregular variance $\hat{\sigma}_I^2$, that was increased by a factor of 20 to make it similar to the sampling error variance. This was done in order to increase the differences between the trend and the seasonally adjusted estimators, and to test the performance of the PB method when applied to an even more variable series. The variances of the primary population model are, $\sigma_I^2 = 2.01,\ \sigma_R^2 = 0.0024,\ \sigma_S^2 = 0.0016$. Denote by $Y_{tr}$ $(r=1...10,000, t=1...84)$ the primary population series. Summing, $y_{tr}^{(j)} = Y_{tr} + e_{tr}^{(j)}$, *j*=1…8 yields 10,000 sets of 8 panel estimates.

*Phase B- Computations for Primary Series*

The computations at this stage were carried out for getting close approximations to the true PMSE of the trend and seasonally adjusted predictors. For each set of panel estimates we re-estimated the sampling error autocorrelations and then solved the Yule-Walker equations for estimating the AR(15) coefficients. Table 3 shows the means and standard deviations of the estimated coefficients over the 10,000 simulated series. Notice that all the coefficients are slightly underestimated by the use of the Yule-Walker estimators, but this is accounted for by the PB method (see below).

**Table 3. AR(15) Model Coefficients of EP-DC Sampling Errors and Mean Estimates and Standard Deviations (SD) over 10,000 Simulated Series.**

| Lag | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| True Coefficients | -0.588 | -0.086 | 0.012 | 0.165 | -0.127 | -0.005 | -0.025 | 0.048 |
| Mean Estimates | -0.587 | -0.082 | 0.010 | 0.152 | -0.118 | -0.004 | -0.022 | 0.042 |
| SD of Estimates | 0.042 | 0.047 | 0.047 | 0.047 | 0.047 | 0.046 | 0.046 | 0.046 |
| Lag | 9 | 10 | 11 | 12 | 13 | 14 | 15 | |
| True Coefficients | -0.086 | 0.022 | -0.089 | -0.072 | 0.022 | 0.029 | -0.026 | |
| Mean Estimates | -0.076 | 0.019 | -0.075 | -0.057 | 0.019 | 0.024 | -0.019 | |
| SD of Estimates | 0.045 | 0.045 | 0.044 | 0.044 | 0.043 | 0.043 | 0.037 | |

Next we computed for each month *t* the mean of the panel series $y_{tr}^{(j)}$, yielding primary 'observed series' $y_{tr} = Y_{tr} + e_{tr} \quad (t = 1...84,\ r = 1...10,000)$. The three variances underlying the population model (4.4) have been estimated for each series by fitting the model to the series $y_{tr} \quad (t = 1...84)$, fixing the $AR(15)$ model coefficients at their estimated values. The variances were estimated by MLE, using the 'prediction error decomposition' for forming the likelihood (Harvey, 1989) and a quasi-Newton algorithm for the maximization process (same procedure as used by the BLS for the real series).

The computations in this study focus on the last time point, *n=84*. Let $\alpha_{84r}$ define the true component value of interest {$L_{84r}$ or ($L_{84r} + \mathrm{I}_{84r}$)} for time *n=84*, as generated for primary series *r* (with hyper-parameters $\lambda = \hat{\lambda}$). Denote by $\alpha_{84r}(\hat{\lambda}_r)$ the corresponding empirical predictor obtained by application of the Kalman filter with hyper-parameter estimates $\hat{\lambda}_r$. The true PMSE is approximated as,

$$MSE_{84} = \sum_{r=1}^{10,000} [\alpha_{84r}(\hat{\lambda}_r) - \alpha_{84r}]^2 / 10,000 \qquad (4.5)$$

*Phase C- Generation of Bootstrap Series and Computations*

At this phase we selected at random 500 primary series from the 10,000 series and applied the following steps for each of the sampled series.

1- Generate 500 bootstrap series of stream sampling errors and population values using the hyper-parameters estimated in Phase B,

2- Estimate the AR(15) sampling error model coefficients and the population model variances for each of the bootstrap series,

3- Estimate the PMSE using the equations (3.1)-(3.2).

The procedures used for generating the bootstrap series in Step 1 and for estimating the hyper-parameters in Step 2 are the same as used for the primary series as described under Phase B. This process was repeated for each of the selected series, yielding 500 estimates of the true PMSE (4.5) computed in Phase B.

### 4.2.3 Results of Simulation Study

Table 4 shows the true root PMSE (R-PMSE) of the trend and seasonally adjusted predictors for time *t=84* (Equation 4.5), and the bias and root mean square error (RMSE) of the PB PMSE estimators $M\hat{S}E_{84}$ (Equation 3.1). Also shown are the bias and RMSE of the *naive* estimator $P_{84}(\hat{\lambda})$, the estimator $[2P_{84}(\hat{\lambda})-\overline{P}_{84}^{bs}]$ of the contribution to the PMSE resulting from 'filter uncertainty' (first component of 2.3, Equation C6 in Appendix C) and the estimator $M\hat{S}E_{84,p}^{bs}$ of the contribution to the PMSE resulting from 'parameter uncertainty' (second component of 2.3, Equation C4 in Appendix C).

**Table 4. Bias and Root MSE ($\times$1000) of Estimators of PMSE of Trend and Seasonally Adjusted Predictors for time t=84. 500$\times$500 replications**

| | Trend | | | | Seasonally Adjusted | | | |
|---|---|---|---|---|---|---|---|---|
| R-PMSE | 1.481 | | | | 1.581 | | | |
| Estimate | $P_{84}(\hat{\lambda})$ | $M\hat{S}E_{84,p}^{bs}$ | $2P_{84}(\hat{\lambda})-\overline{P}_{84}^{bs}$ | $M\hat{S}E_{84}$ | $P_{84}(\hat{\lambda})$ | $M\hat{S}E_{84,p}^{bs}$ | $2P_{84}(\hat{\lambda})-\overline{P}_{84}^{bs}$ | $M\hat{S}E_{84}$ |
| Bias | -157 | -85 | -71 | -4 | -99 | -42 | -42 | 12 |
| RMSE | 235 | 188 | 216 | 196 | 145 | 121 | 119 | 117 |

The results displayed in Table 4 illustrate again the very good performance of the PB method in eliminating the large negative bias of the naive PMSE estimator. This result is particularly encouraging considering that we used for this study a complex model with 18 unknown hyper-parameters estimated by a three-step procedure. The estimators $2P_{84}(\hat{\lambda}) - \bar{P}_{84}^{bs}$ and $M\hat{S}E_{84,p}^{bs}$ also reduce the bias but each of these estimators only accounts for one component of the PMSE. Notice again that correcting the bias of the na?ve estimator does not necessarily imply a similar relative reduction in the RMSE. Thus, while the naive estimator has the largest RMSE in both parts of the table, the RMSE of the other three estimators are quite similar. As mentioned before, the addition of bias correction terms often increases the variance.

## 5. CONCLUDING REMARKS

The bootstrap method proposed in this paper for estimating the PMSE has four important advantages. First and foremost, it yields estimators with bias of correct order. Second, it does not require extra assumptions regarding the distribution of the hyper-parameters or their estimators, beyond the mild assumptions on the moments of the estimators. Third, it is not restricted to MLE or REML hyper-parameter estimators. Fourth, it is very general and can be used for a variety of models and prediction problems. We mention again that the other methods proposed in the literature for estimating the PMSE are restricted to MLE or REML hyper-parameter estimators and they either require the specification of prior distributions for the hyper-parameters, or that they assume that the hyper-parameters estimators have approximately a normal distribution. As illustrated by Figure 1 and 2, this assumption may not hold in practice.

The state-space model considered in this study is linear but in view of the mild assumptions underlying the use of the method, it can be surmised that with appropriate modifications the method could be applied also to nonlinear state-space models. Durbin and Koopman (2000) consider the fitting of such models from both the frequentist and the Bayesian perspectives and propose the use of simulations for predicting the state vector and computing the PMSE. Interestingly, the authors comment that “A weakness of the classical approach is that it does not automatically allow for the effect on estimates of variance of estimation errors in estimating the hyper-parameters". In the discussion of this paper, A. Harvey makes a similar comment. Incorporating the proposed method for PMSE estimation into the simulation method underlying this approach seems natural but it requires further theoretical and empirical investigation.

## APPENDIX A: *The Kalman Filter and the Innovation Form Representation*

The Kalman filter consists of a set of recursive equations that are used for updating the predictors of current and future state vectors and the corresponding prediction error variance-covariance (PE-VC) matrices, every time that new data become available. The filter assumes known hyper-parameters. Below we consider the model defined by (1.1) and (1.2) with known hyper-parameters $\lambda$.

Let $u_{t-1}(\lambda)$ define the best linear unbiased predictor (BLUP) of $u_{t-1}$ based on observations $y_{(t-1)} = \langle y_1 ... y_{t-1} \rangle$ and denote by $\tilde{P}_{t-1} = E\{[u_{t-1}(\lambda) - u_{t-1}][u_{t-1}(\lambda) - u_{t-1}]'\}$ the corresponding PE-VC matrix. The BLUP of $u_t$ at time $(t-1)$ is $u_{t|t-1}(\lambda) = G_t u_{t-1}(\lambda)$, with PE-VC $\tilde{P}_{t|t-1} = G_t \tilde{P}_{t-1} G_t' + Q_t$. When a new observation $y_t$ becomes available, the state predictor and the PE-VC are updated as,

$$u_t(\lambda) = K_t y_t + (I - K_t Z_t) u_{t|t-1}(\lambda) \quad ; \quad \tilde{P}_t = \tilde{P}_{t|t-1}(I - Z_t' F_t^{-1} Z_t \tilde{P}_{t|t-1}) \tag{A1}$$

where $K_t = \tilde{P}_{t|t-1} Z_t' F_t^{-1}$, and $F_t = Z_t \tilde{P}_{t|t-1} Z_t' + \Sigma_t$ is the PE-VC of the innovations $v_t = y_t - \hat{y}_{t|t-1} = y_t - Z_t u_{t|t-1}(\lambda)$. By (A1) and the relationship $u_{t+1|t}(\lambda) = G_{t+1} u_t(\lambda)$,

$$y_t = Z_t u_{t|t-1}(\lambda) + v_t \quad ; \quad u_{t+1|t}(\lambda) = G_{t+1} u_{t|t-1}(\lambda) + G_{t+1} K_t v_t \tag{A2}$$

The two equations in (A2) define the *innovation form representation* of the state-space model (1.1)-(1.2). The standardized innovations used for the NPB method are,

$$\tilde{v}_t = F_t^{-1/2} v_t \tag{A3}$$

## APPENDIX B: ***Rate of Convergence*** of $E[\alpha_t(\hat{\lambda})-\alpha_t(\lambda)]^2$

Below we define conditions under which $E[\alpha_t(\hat{\lambda})-\alpha_t(\lambda)]^2$ (the second component of the PMSE in 2.3) is of order $O(1/n)$, assuming for convenience that the matrices $Z_t$ and $G_t$ are known for all $t$. We first consider the case where the state vector of interest corresponds to the last time point with observations ($t=n$). Suppose that,

1- $\alpha_n(\hat{\lambda})-\alpha_n(\lambda)=[\partial\alpha_n(\lambda)/\partial\lambda]'(\hat{\lambda}-\lambda)+O_p(1/n)$ and that the $O_p(1/n)$ term is negligible compared to the first term.

2- $E[(\hat{\lambda}-\lambda)(\hat{\lambda}-\lambda)' | \partial\alpha_n(\lambda)/\partial\lambda]=V(\lambda)/n+o_p(1/n)$, where $V(\lambda)/n=E[(\hat{\lambda}-\lambda)(\hat{\lambda}-\lambda)']$

3- $\lim_{n\to\infty}\sum_{j=1}^{n-1}M_{n,j}M'_{n,j}=\Psi<\infty$, where $M_{n,j}=R_n\times R_{n-1}\times\ldots\times R_{n-j+1}$ $(j=1\ldots n)$; $R_t=(I-K_tZ_t)G_t$

4- The matrices $[\partial K_t/\partial\lambda]F_t[\partial K_t/\partial\lambda]'$ are bounded for all $t$.

Conditions 1 and 2 are the same as in Ansley and Kohn (1986). Condition 1 is mild while Condition 2 will be true if $(\hat{\lambda}-\lambda)$ is approximately independent of $\alpha_n(\lambda)$. Ansley and Kohn establish asymptotic independence between the two terms for ARMA models by showing that $\hat{\lambda}$ depends approximately equally on all the observations whereas $\alpha_n(\lambda)$ depends mostly on the observations around $n$, with the weights assigned to the other observations decreasing exponentially as the distance from $n$ increases. As shown next, this argument applies to more general state-space models satisfying Condition 3, which itself is not binding (see below).

To see this, rewrite the left-hand side equation in (A1) as $u_t(\lambda)=K_ty_t+R_tu_{t-1}(\lambda)$. Repeated substitutions of this equation yields the relationship,

$$u_n(\lambda)=K_ny_n+M_{n,1}K_{n-1}y_{n-1}+M_{n,2}K_{n-2}y_{n-2}+\ldots+M_{n,n-1}K_1y_1+M_{n,n}\hat{u}_0 \qquad \text{(B1)}$$

where $\hat{u}_0$ defines the initial state predictor. By Condition 3, $M_{n,j}\underset{n\to\infty,\,j\to n}{\to}0$, illustrating the decrease in the weights assigned to past values as the distance from $n$ increases. Thus, Condition 3 is a natural requirement for any prediction rule applied to a non-trivial time series model. Note that if the model is time invariant in the sense that the matrices $Z_t, G_t, \Sigma_t, Q_t$ are fixed over time, the Kalman filter converges to a steady state with V-C matrices $\tilde{P}_{t|t-1} = \tilde{P},\ F_t = F$ (Harvey, 1989). In the steady state, $K_t = K,\ R_t = R$, illustrating that in this case the weights in (B1) decay exponentially.

Comment: We computed the empirical correlations between the MLE of the three variances of the population model (4.4) and the trend and the seasonally adjusted estimators for time $n = 84$, using the 10,000 primary series generated for the simulation study. The largest correlation found was $Corr(L_{84}, \hat{\sigma}_s^2) = 0.018$, illustrating the approximate independence between the state vector and the MLE of the hyper-parameters under this model.

**Proposition 1:** Under the conditions 1-4, $E[\alpha_n(\hat{\lambda}) - \alpha_n(\lambda)]^2 = O(1/n)$.

**Proof:** By Conditions 1 and 2 it is sufficient to show that $E\{[\partial\alpha_n(\lambda)/\partial\lambda][\partial\alpha_n(\lambda)/\partial\lambda]'\}$ is bounded, and since $\alpha_n(\lambda) = l_n' u_n(\lambda)$, the problem reduces to showing that $E\{[\partial u_n(\lambda)/\partial\lambda_i][\partial u_n(\lambda)/\partial\lambda_j]'\}$ is bounded for all $1 \le i, j \le m$ where $m = \dim(\lambda)$. Write the left hand side equation of (A1) for *t=n* as, $u_n(\lambda) = G_n u_{n-1}(\lambda) + K_n v_n$ where $v_n$ is defined below (A1). Differentiating both sides with respect to $\lambda_i$ yields,

$$\partial u_n(\lambda)/\partial\lambda_i = R_n \partial u_{n-1}(\lambda)/\partial\lambda_i + [\partial K_n/\partial\lambda_i]v_n \qquad \text{(B2)}$$

The two terms in the right-hand side of (B2) are uncorrelated. To see this, write $v_n = y_n - Z_n G_n u_{n-1}(\lambda) = Z_n\eta_n + \varepsilon_n + Z_n G_n[u_{n-1} - u_{n-1}(\lambda)]$ and note that $u_{n-1}(\lambda) = E(u_{n-1} \mid y_{(n-1)})$, implying that $E(v_n \mid y_{(n-1)}) = 0$ and $E\{[\partial u_{n-1}(\lambda)/\partial\lambda_i]v_n \mid y_{(n-1)}\} = 0$. It follows therefore that,

$$\begin{aligned} E\{[\partial u_n(\lambda)/\partial\lambda_i][\partial u_n(\lambda)/\partial\lambda_j]'\} = R_n E\{[\partial u_{n-1}(\lambda)/\partial\lambda_i][\partial u_{n-1}(\lambda)/\partial\lambda_j]'\}R_n' \\ + (\partial K_n/\partial\lambda_i)F_n(\partial K_n/\partial\lambda_j)' \end{aligned} \qquad \text{(B3)}$$

Repeated substitutions in (B3) and assuming that the vector $\hat{u}_0$ used for the initialization of the filter does not depend on $\lambda$ yields the relationship,

$$E\{[\partial u_n(\lambda)/\partial\lambda_i][\partial u_n(\lambda)/\partial\lambda_j]'\} = (\partial K_n/\partial\lambda_i)F_n(\partial K_n/\partial\lambda_j)' + \sum_{j=1}^{n-1} M_{n,j}(\partial K_{n-j}/\partial\lambda_i)F_{n-j}(\partial K_{n-j}/\partial\lambda_j)'M'_{n,j} \tag{B4}$$

The matrices $(\partial K_t/\partial\lambda)F_t(\partial K_t/\partial\lambda)'$ are symmetric and positive semi-definite and by Condition 4 they are bounded, so that the proof is completed by means of Condition 3.

Comment: By assuming time invariant matrices $Z_t = Z$, $G_t = G$, 'strong consistency' of $\hat{\lambda}$ ($\hat{\lambda} \to \lambda$ *a.s.* as $n \to \infty$) and some other regularity conditions, Watanabe(1985) shows that $\underset{n\to\infty}{p\lim}[\alpha_n(\hat{\lambda}) - \alpha_n(\lambda)] = 0$, employing a similar representation to (B1).

The analysis so far is restricted to the case where the state vector of interest corresponds to the last time point with observations, $t = n$. When $t < n$, the optimal state predictor and the corresponding PE-VC are obtained by application of an appropriate smoothing algorithm, and it is easy to verify that Proposition 1 holds in this case as well. One way to see this is by noticing that the smoothed state predictor for time $t$ can be computed by augmenting the state vectors at times $(t+1)...n$ by the vector $u_t$ and applying the Kalman filter to the augmented model. The smoothed predictor of $u_t$ is then the (Kalman filter) predictor obtained for $u_t$ at the last time point, $n$, for which the proposition was shown to hold.

Finally, consider the case $t > n$ and let $t = n + r,\ r > 0$. In this case, $u_{t|n}(\lambda) = G_{n+r}G_{n+r-1}...G_{n+1}u_n(\lambda) = B_{n,r}u_n(\lambda)$, $u_{t|n}(\hat{\lambda}) = B_{n,r}u_n(\hat{\lambda})$, such that $\alpha_{t|n}(\hat{\lambda}) - \alpha_{t|n}(\lambda)$ $= l_t' B_{n,r}[u_n(\hat{\lambda}) - u_n(\lambda)]$. For Proposition 1 to hold in this case it is sufficient that the matrix $B_{n,r}$ is bounded, which would be the case if $r$ is fixed. If, however, $r$ is allowed to increase with $n$, it is necessary to require that $\lim_{n\to\infty, r\to\infty} B_{n,r} = B < \infty$ for the proposition to hold. This requirement is satisfied when the state vectors are stationary or follow a random walk model.

**APPENDIX C:** ***Proof of Theorem***

By (2.3) the true PMSE is,

$$\begin{aligned} MSE_t &= E[\alpha_t(\hat{\lambda})-\alpha_t]^2 = E[\alpha_t(\lambda)-\alpha_t]^2 + E[\alpha_t(\hat{\lambda})-\alpha_t(\lambda)]^2 \\ &= P_t(\lambda) + E[\alpha_t(\hat{\lambda})-\alpha_t(\lambda)]^2 \end{aligned} \tag{C1}$$

Denote by $E_*$ the expectation with respect to the bootstrap distribution (the distribution over all possible bootstrap series generated with hyper-parameter $\hat{\lambda}$). Then, analogously to (C1),

$$\begin{aligned} E_*[\alpha_t^b(\hat{\lambda}^b)-\alpha_t^b]^2 &= E_*[\alpha_t^b(\hat{\lambda})-\alpha_t^b]^2 + E_*[\alpha_t^b(\hat{\lambda}^b)-\alpha_t^b(\hat{\lambda})]^2 \\ &= P_t(\hat{\lambda}) + E_*[\alpha_t^b(\hat{\lambda}^b)-\alpha_t^b(\hat{\lambda})]^2 \end{aligned} \tag{C2}$$

where $\alpha_t^b = l_t' u_t^b$; $u_t^b$ is the true state vector of bootstrap series *b* at time *t* and $\alpha_t^b(\hat{\lambda})$ and $\alpha_t^b(\hat{\lambda}^b)$ are the predictors of $\alpha_t^b$ using the 'true' parameter $\hat{\lambda}$ and the estimator $\hat{\lambda}^b$ respectively.

Under Condition $\mathrm{II}$ of the theorem, $E[\alpha_t(\hat{\lambda})-\alpha_t(\lambda)]^2 = O(1/n)$ (see also Appendix B). Hence, using results from *Hall and Martin (1988)*,

$$E\{[\alpha_t(\hat{\lambda})-\alpha_t(\lambda)]^2 - E_*[\alpha_t^b(\hat{\lambda}^b)-\alpha_t^b(\hat{\lambda})]^2\} = O(1/n^2) \tag{C3}$$

so that we can estimate,

$$\hat{E}[\alpha_t(\hat{\lambda})-\alpha_t(\lambda)]^2 \hat{=} \frac{1}{B}\sum_{b=1}^{B}[\alpha_t^b(\hat{\lambda}^b)-\alpha_t^b(\hat{\lambda})]^2 = MSE_{t,p}^{bs} \tag{C4}$$

Similarly, by Condition $\mathrm{I}$, $E[P_t(\hat{\lambda})-P_t(\lambda)] = O(1/n)$. Hence,

$$E\{[P_t(\hat{\lambda})-P_t(\lambda)] - E_*[P_t(\hat{\lambda}^b)-P_t(\hat{\lambda})]\} = O(1/n^2) \tag{C5}$$

implying,

$$\hat{P}_t(\lambda) \hat{=} P_t(\hat{\lambda}) - \frac{1}{B}\sum_{b=1}^{B}[P_t(\hat{\lambda}^b)-P_t(\hat{\lambda})] = 2P_t(\hat{\lambda}) - \bar{P}_t^{bs} \tag{C6}$$

It follows from (C1), (C4) and (C6) that for B sufficiently large the estimator $M\hat{S}E_t$ defined by (3.1) has bias of order $O(1/n^2)$. QED

The estimator defined by (3.3) is obtained by first replacing $MSE_{t,p}^{bs}$ by $E_*[\alpha_t^b(\hat{\lambda}^b) - \alpha_t^b(\hat{\lambda})]^2$ and then replacing $P_t(\hat{\lambda}) + E_*[\alpha_t^b(\hat{\lambda}^b) - \alpha_t^b(\hat{\lambda})]^2 = E_*[\alpha_t^b(\hat{\lambda}^b) - \alpha_t^b]^2$ [follows from (C2)] by $MSE_t^{bs} = \frac{1}{B}\sum_{b=1}^{B}[\alpha_t^b(\hat{\lambda}^b) - \alpha_t^b]^2$ .

## REFERENCES


Ansley, C. F. and Kohn, R. (1986) Prediction mean squared error for state space models with estimated parameters. *Biometrika* 73, 467-473.

Box, G. E. P. and Jenkins, G. M. (1976) *Time Series Analysis: Forecasting and Control* (2nd edition). San Francisco: Holden-Day.

Datta, G.S. Lahiri, P. and Lu, K.L. (1999) Hierarchical Bayes estimation of unemployment rates for the states of the U.S. *Journal of the American Statistical Association* 94, 1074-1082.

De Jong, P. (1989) Smoothing and Interpolation with the state-space Model. *Journal of the American Statistical Association* 84, 1085-1088.

Durbin, J. and Koopman, S.J. (2000) Time series analysis of non-Gaussian observations based on state-space models from both classical and Bayesian perspectives (with discussion). *Journal of the Royal Statistical Society Series B* 62, 3-56.

Hamilton, J. D. (1986) A standard error for the estimated state vector of a state space model. *Journal of Econometrics* 33, 387-397.

Hall, P. and Martin, M. A. (1988) On bootstrap resampling and iteration. *Biometrika* 75, 661-671.

Harvey, A. C. (1989) *Forecasting Structural Time Series with the Kalman Filter*. Cambridge: Cambridge University Press.

Koopman, S. J. Harvey, A. C. Doornik, J. A. and Shephard, N. (1995) *Stamp 5.0-Structural Time Series Analyser*, *Modeller and Predictor* : London: Chapman & Hall.

Quenneville, B. and Singh, A. C. (2000) Bayesian prediction mean square error for state space models with estimated parameters. *Journal of time series Analysis* 21, 219-236.

Pfeffermann, D. (2002) Small area estimation-new developments and directions. *International Statistical Review* 70, 125-143.

Pfeffermann, D. Feder, M. and Signorelli, D. (1998) Estimation of autocorrelations of survey errors with application to trend estimation in small areas. *Journal of Business and Economic Statistics* 16, 339-348.

Sallas, W. H. and Harville, D. A. (1981) Best linear recursive estimation for mixed linear models. *Journal of the American Statistical Association* 76, 860-869.

Shao, J and Tu, D. (1995) *The Jackknife and the Bootstrap*. New York: Springer-verlag

Singh, A.C. Stukel, D. M. and Pfeffermann, D. (1998) Bayesian versus frequentist measures of error in small area estimation. *Journal of the Royal Statistical Society Series B* 60, 377-396.

Stoffer, D.S. and Wall, K.D. (1991) Bootstrapping state-space models: Gaussian maximum likelihood estimation and the Kalman filter. *Journal of the American Statistical Association* 86, 1024-1033.

Stoffer, D.S. and Wall, K.D. (2002) Resampling in state space models. Paper presented at the Colloquium on ‘state space and unobserved components models’ in honour of J. Durbin, Website http://www.tinbergen.nl/~durbinconf/

Tiller, R. B. (1992) Time series modeling of sample survey data from the U.S. Current Population Survey. *Journal of Official Statistics* 8, 149-166.

Tsimikas J. and Ledolter, J. (1994) REML and best linear unbiased prediction in state space models. *Communication in Statistics, Theory and Methods* 23, 2253-2268.

Watanabe, N. (1985) Note on the Kalman filter with estimated parameters. *Journal of Time Series Analysis* 6, 269-278.

**Figure 1. Distribution of $0.5^a \log(\hat{q})$ Under Gaussian RWN Model (4.1) with q=0.25. 5000 Series, T=40**

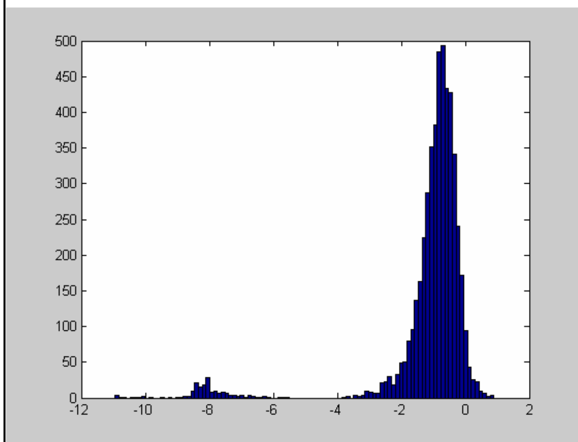


**Figure 2. Distribution of $\log(\hat{\sigma}_R)$ under BSM (4.4) with $\sigma_R = 0.049$. 10,000 series, T=84**

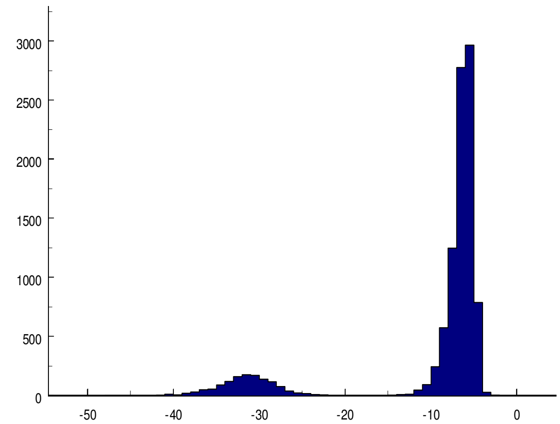


**Figure 3. Distribution of error terms $\varepsilon_t$ when $\varepsilon_t = v_t - (4/3)$, $v_t \sim G(16/9, 3/4)$**

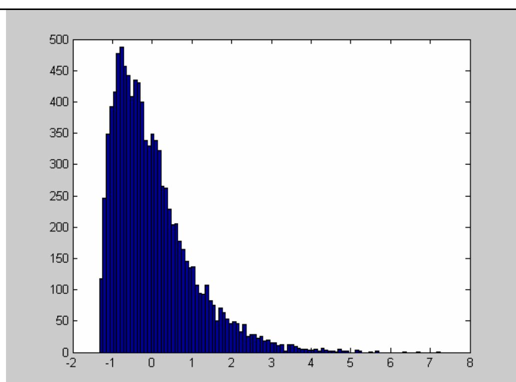


**Figure 4. EP-DC series and Estimated Trend Under the Model. 1998-2003**